

Data Mining par la pratique

Cours Pratique de 3 jours - 21h

Réf : DMP - Prix 2024 : 2 280€ HT

Le Data Mining consiste à découvrir des modèles, des correspondances, des motifs dans un ensemble de données numériques ou qualitatives. Cette activité s'appuie sur une trousse à outils algorithmiques qui vous seront présentés dans ce cours. La démarche Data Mining sera illustrée sur plusieurs projets, en utilisant R.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

Comprendre l'intérêt de l'approche Data Mining

Traduire une problématique et savoir y répondre

Connaître les principales méthodes de Data Mining

Identifier et savoir utiliser les outils de Data Mining

Poser une problématique de Data Mining et rechercher la méthode appropriée

Être en capacité de restituer les résultats

LE PROGRAMME

dernière mise à jour : 06/2021

1) Le projet de Data Mining

- La problématique du Data Scientist : de la donnée à l'information.
- Le vocabulaire et les concepts.
- L'exploration descriptive du jeu de données.
- Les métadonnées pour le suivi du projet de Data Mining.
- Rappels sur le logiciel R.

Travaux pratiques : Utilisation de R. Caractérisation descriptive, définition et saisie des métadonnées d'un jeu de données.

2) Les techniques de l'approche Data Mining

- Méthode basée sur la classification : identification de groupes d'individus statistiques.
- Méthode par association : mise en évidence d'une cause et d'une conséquence.
- Apports du Data Mining pour le traitement des grandes volumétries de données.
- Méthode de l'estimation : complément d'un effectif ou d'une fréquence d'un jeu de données.
- Méthode de segmentation : définition de critères, extension de la méthode de classification et principe des k-means.
- Méthode de prévision : importance de la temporalité et des hypothèses.

Travaux pratiques : Appréhender les différentes méthodes en fonction des besoins exprimés.

3) Les outils statistiques

- Méthodes descriptives : corrélation, classification, réseaux de Kohonen, règles d'association.
- Méthodes prédictives : régression, arbres de décision, réseaux de neurones, les K plus proches voisins.
- Mise en œuvre de classification par k-means et CAH (Classification Ascendante Hiérarchique).
- Principe des méthodes supervisées.

Travaux pratiques : Mise en pratique des différentes méthodes sous R.

PARTICIPANTS

Chargés d'études, responsables de projet analyse de données, responsables Infocentre, Marketing ou Qualité, utilisateurs et gestionnaires métiers de bases de données, futurs Data Scientist.

PRÉREQUIS

Connaissances de base en statistiques ou connaissances équivalentes à celles apportées par le stage "Statistiques descriptives, introduction" (réf. UES).

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...

Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

4) La visualisation des données

- Les objectifs de la visualisation des données.
- Les différents types de représentations de données quantitatives.
- Concevoir des tableaux de bord.

Travaux pratiques : Création d'un tableau de bord avec R en utilisant des données quantitatives. Représentation de données quantitatives et de données qualitatives avec R.

5) Analyse de données qualitatives et textuelles

- Spécificités de la problématique et les alternatives (Analyse Factorielle des Correspondances, table de contingence).
- Présentation d'instanciation, de pattern, de vecteur et d'heuristique.
- Comment utiliser un espace de vecteurs, d'indexation, de scoring.
- Différents types de transformations et traitement d'un document textuel.

Travaux pratiques : Réalisation d'un traitement de données qualitatives et de données textuelles sous R.

LES DATES

CLASSE À DISTANCE

2025 : 10 mars, 28 avr., 09 juil.,
15 oct.

PARIS

2025 : 03 mars, 14 avr., 02 juil.,
08 oct.